Project Abstract

Intentional Vision in Humans and Robots
0433653
Daniel Levin, Kazuhiko Kawamura, Megan Saylor, Mitchell Wilkes
Vanderbilt University

Overall Mission

As computers become increasingly powerful, they will become progressively more and more integrated into the real world. This phenomenon is especially salient for the kind of humanoid robots that are currently being developed to fill real-world functions ranging from household chores to elder-care. Among the challenges these devices pose, perhaps the most difficult is the need for a two-way understanding between the robots and their human users. Not only do humans need to understand robot capabilities and representational states, but robots require the same understanding of humans. This is particularly true if robots are to have productive and flexible interactions with humans, a process that requires a careful alignment of understanding that is dynamic enough to coordinate a complex flow of changing circumstances, beliefs, desires, and intentions. The research proposed here represents an attempt to understand A) how people will construe the representational states of robots, particularly with respect to vision and B) the cognitive and perceptual basis for and effects of this construal.

An important issue underlying this research is that people have a specific set of expectations about how other people think about, and represent things in the world. These expectations are sometimes referred to as an "Intentional Theory of Mind", and they are applied to a wide range of tasks as people attempt to infer others' beliefs and desires based on the behaviors they produce. Previous research by the PI suggests that this intentional theory is sometimes the most readily available means people have of thinking about representations, so it is sometimes misapplied to representation-making systems (such as computers) for which it is not appropriate, and can even lead to mispredictions about human representations. On the other hand, it may be that people have some understanding of the difference between different kinds of representational systems, and that problems occur only when they fail to apply these understandings. Generally, we have been quite surprised to discover how little is known about people's expectations about the inner workings not only of robots, but of computers more generally, despite the large amount of research exploring human-computer interactions.

Progress and preliminary outcomes

The activities funded by the grant have been organized into several specific projects. In the first, we have been exploring people's beliefs about the representations inherent to different kinds of living and mechanical systems. We have therefore been asking children and adults questions derived from basic research on concepts that are designed to assess the degree to which the mental functions of computers and humans are presumed to differ. For example, in an initial experiment on adults we found that subjects are willing to generalize novel mental properties from people to computers, but only if they are nonintentional. To begin understanding the basis for these attributions we have been asking children about the metal, physical and biological properties of humans, computers, and robots. At age three, children successfully differentiate these three systems, and by age 4 they seem to apply subtly different notions of thinking and seeing to the systems. Also interesting is the finding that young children seem comfortable attributing both cognitive and mechanical (but not biological) properties to robots, implying that foundational concepts about living and nonliving things are sufficiently flexible to include anthropomorphic objects such as robots.

In a second project we have been exploring the degree to which concepts such as the ones described above affect people's interpretations of actions, their expectations about how different kinds of systems interpret actions, and how people perform actions. In one series of experiments, we have been asking subjects to demonstrate simple procedures (for example tying a shoe, or completing the Towers of Hanoi problem) for either a human or computer audience. We have found that subjects engage in systematically different movements and looking behavior for the two audiences. For example, subjects looked more at a picture illustrating a human audience than at a computer audience as they demonstrate the actions. On the other hand, subjects provided more visual emphasis for the segments of subactions when demonstrating for a computer audience. In addition, we asked subjects to mark action segment start- and end-points as they would be perceived by a computer or a person. Subjects consistently broke actions into smaller and more numerous segments for computers. Most interesting, this effect was strongest in subjects who responded that computers are poor at understanding human goals and intentions (as indicated in a post-experiment survey). Combined, these experiments show that people have specific expectations about how intentional and noninentional systems perceive action, and our next experiments will begin to explore how the kinds of anthropomorphism inherent to robots affects these expectations. Finally, we are piloting experiments testing the effects of being observed while completing simple actions. These experiments will explore how being watched by a person, by a computer, or by a robot will affect hand movements as subjects touch covers that they have hidden an object under. This experiment will make use of a remote-control robot head adapted for this project. We hope to test the effects of different realistic (and nonrealistic) robotic eye movements on observed actions.

In a third project, we have been exploring how people move when they demonstrate these actions. Initial work has focused on tracking hand and head movements while people demonstrate actions using both video analyses, and magnetic tracking of hand movements through 3-D space. We have successfully developed a color- and texture-based video tracking system that processes action demonstration videos to locate the actor's hands, their head direction, and the objects they are handling. These will be used to correlate detected motions with actions segments provided by human subjects. In addition, we have been analyzing directly-recorded 3-D positions for breakpoints as defined by directional changes, and have applied an "Isomap" analysis to the resulting segments to provide an automatic grouping of actions. We plan to use these analyses as predictors for action breakpoints, and as a means of objectively distinguishing action demonstrations for different audiences, and purposes.

Broader Impact

This research will not only enrich the existing collaborations between the cognitive science and engineering communities at Vanderbilt, but it will also have a broader educational impact. Testing these ideas in the context of a humanoid robot will also provide a compelling context for both graduate and undergraduate students to consider basic questions of representation and mind, and Vanderbilt undergraduates have been playing an important role in this research. In addition, the project will be featured in a video created by the Learning Sciences Institute here at Vanderbilt. This video covers a range of projects in the learning sciences, and based on the filming we have already done for our short segment, we plan to create a longer video on this project, intended for a broader nonacademic audience.